ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ
ΣΥΣΤΗΜΑΤΩΝ

ΙΝΣΤΙΤΟΥΤΟ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

Προτεινόμενα Θέματα Διπλωματικής Εργασίας

Καθηγητής Ευστάθιος Σταματάτος

A. Cross-genre Authorship Attribution Using Representation Learning

Abstract:

Authorship attribution aims to reveal the identity of the person who wrote a text. In challenging cross-genre conditions, the samples of known authorship and the samples of disputed authorship belong to different genres (e.g., emails, essays, interviews).

The purpose of this thesis is to study recent representation learning approaches[1][2] that use deep learning to extract style-related embeddings for input documents and explore who these representations can be used to optimize performance in cross-genre authorship attribution conditions. Experiments on existing datasets will be performed comparing representation learning methods with traditional n-gram based approaches.

---

[1] Rafael A. Rivera-Soto, Olivia Elizabeth Miano, Juanita Ordonez, Barry Y. Chen, Aleem Khan, Marcus Bishop, and Nicholas Andrews. 2021. Learning Universal Authorship Representations. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 913–919.

[2] B. Boenninghoff, S. Hessler, D. Kolossa and R. Nickel, "Explainable Authorship Verification in Social Media via Attention-based Similarity Learning," in 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 2019 pp. 36-45. (doi: 10.1109/BigData47090.2019.9005650)